CS 6200 Information Retrieval
Fall 2019

Syllabus & Schedule [new/updated sections will be colored purple]

This course provides an overview of the important issues in information retrieval, with a focus on how these issues affect the design and implementation of search engine software. The course will emphasize the technology used in Web search engines, and the concepts underlying search applications.

Prerequisites
The course will require the use of basic probability and linear algebra. If you do not have this background, you must be willing to learn it on your own. You will be expected to complete your programming assignments in Python. Learn Python if you need to do so.

Instructor: Raman Chandrasekar ("Chandra") , r.chandrasekar@northeastern.edu
Class meetings: Mondays, 6:15pm – 9:15pm, 225 Terry Ave N, Room 307
Office Hours: Mondays, 5:00pm – 6:00pm, 225 Terry Ave N, Room 111

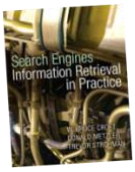
TA:
Kartik Dave dave.ka@husky.neu.edu Office Hours: To be announced

Please do make full use of TA Office Hours.

Class Resources
Lecture slides will be posted on Piazza, after each lecture. Assignments/Quizzes will be posted on Blackboard, with instructions and due dates, as they are assigned. Look out for announcements on Blackboard and/or Piazza.

Piazza signup link: https://piazza.com/northeastern/fall2019/cs6200seattle Access Code: SeattleIR
Piazza discussion board: https://piazza.com/northeastern/fall2019/cs6200seattle/home

Textbook
The textbook for this course is:



Search Engines: Information Retrieval in Practice
Bruce Croft, Donald Metzler and Trevor Strohman,
Addison-Wesley. 2010. ISBN: 978-0-136-07224-9
[Abbreviated as CMS]

Other texts that may be useful:

- Introduction to Information Retrieval, Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Cambridge University Press. 2008. ISBN: 0521865719. [Abbreviated as MRS]
  HTML/PDF versions available at: https://nlp.stanford.edu/IR-book/information-retrieval-book.html
- Information Retrieval: Implementing and Evaluating Search Engines, Stefan Büttcher, Charles LA Clarke, and Gordon V. Cormack, MIT Press. 2010. ISBN: 978-0-262-02651-2 [Abbreviated as BCC]

Course Grading, Policies and Expectations

- *Final Exam*
  There will be one examination. This will constitute 20% of the course grade. The exam is closed book, closed notes, except for a one-page 'cheat-sheet' (hand-written by you).

- *Homework Assignments*
  There will be 5 homework assignments, for a maximum of 70% of the course grade. Four of the assignments will involve the implementation of components of a search engine. The outputs required for these will typically include code, output from running the code on specific datasets, and a brief note on the implementation and results. These assignments will be equally weighted at 15% each. One other assignment will involve a study, report and presentation and will be graded for a maximum of 10%.

  The goal of the assignments is for students to *learn by doing*, i.e. to understand the assignment problems and work through the solutions individually. Students may discuss the problems with others (classmates, TA, instructor), but each student will be expected to work on his/her code, solutions and write-up without reading or copying others' solutions. See schedule for due date details.

- *Weekly Quizzes*
  There will be a quiz (almost) each week, *graded for participation*, which will together contribute a maximum of 10% of the course grade. The quizzes will be released at the end of each lecture and will be due by 9am of the next Monday. Each quiz will be for 1% of the course grade, and the best 10 scores on these quizzes will contribute to the total quiz score for each student, up to a maximum of 10%.

Late policy

Assignments are due on the announced due date at 9:00 a.m. Late assignments will be penalized 10% for each calendar day of delay. As an example, if a perfect assignment is turned in 3 days late, it will be marked at 70% instead of 100%. However, you have 3 slip days which may be used for delayed submission of assignments. Slip days will be computed in days and rounded up to the nearest day. If you know in advance that your submission is likely to be delayed, please let the instructor know. Note that the project/some assignments may not allow for delayed submission. ***Quizzes and assignments will not be accepted after the solutions have been handed out or discussed in class.***

Academic Integrity: ***All work submitted for credit must be your own.*** You may discuss the quizzes, homework problems or projects with your classmates, the TAs, and/or the instructor, but you must write up your own code and solutions. You must acknowledge the people with whom you discussed your work, and any written sources used. The University's policy on academic integrity is available in the Student Handbook (and at http://www.northeastern.edu/osccr/academic-integrity-policy/).

Title IX: Title IX of the Education Amendments of 1972 protects individuals from sex or gender-based discrimination, including discrimination based on gender-identity, in educational programs and activities that receive federal financial assistance. Northeastern's Title IX Policy prohibits Prohibited Offenses, which are defined as sexual harassment, sexual assault, relationship or domestic violence, and stalking. The Title IX Policy applies to the entire community, including students, faculty and staff of all gender identities. If you or someone you know has been harassed or assaulted, you can find the appropriate resources here: www.northeastern.edu/ouec

Accommodations for Students with Disabilities*:* If you have a disability-related need for reasonable academic accommodations in this course and have not yet met with a Disability Specialist, please visit www.northeastern.edu/drc and follow the outlined procedure to request services. If the Disability Resource Center has formally approved you for an academic accommodation in this class, please present the instructor with your "Professor Notification Letter" during the first week of the semester, so that we can address your specific needs as early as possible.

CS 6200 Information Retrieval: Schedule

Fall 2019, Seattle

Week #	Date	Topic	Readings	Deadlines
1	9/9/2019	1: Overview of Information Retrieval (IR) 2: Search Engine Architecture	CMS Chapters 1 & 2, Brin & Page	
2	9/16/2019	3: Acquiring Data	CMS Chapter 3	Assignment1 released
3	9/23/2019	4: Transforming Data	CMS Chapter 4	
4	9/30/2019	5: Indexing & Ranking	CMS Chapter 5, Sections 5.1 – 5.6	Assignment1 due
5	10/ 7/2019	6: Query Processing	CMS Sections 5.7, & 4.5; Advanced Operators	Assignment2 released
6	10/14/2019	7: Query Refinement	CMS Chapter 6, Sections 6.1, 6.2	
7	10/21/2019	8: Retrieval Models – 1	CMS Chapter 7, Sections 7.1, 7.2	Assignment2 due Assignment3 released
8	10/28/2019	9: Retrieval Models – 2	CMS Chapter 7, Sections 7.3, 7.5 --7.7	
9	11/ 4/2019	10: Evaluating Search Engines	CMS Chapter 8, except for Section 8.6	Assignment3 due Assignment4 released
10	11/11/2019	Lecture 10 continued, plus 11: User Experience	CMS Chapter 6, Sections 6.3, 6.4; Dziadosz & Chandrasekar	
11	11/18/2019	Lecture 11: User Experience, continued	CMS Chapter 6, Sections 6.3, 6.4; Dziadosz & Chandrasekar	Assignment4 due Assignment5 released
12	11/25/2019	Guest Lecture/ 12: Specialized Search	CMS Chapter 10 and Chapter 11 – Sections 11.4, 11.5, 11.6	
13	12/ 2/2019	Student Presentations		Assignment5 due
14	12/9/2019	12: Specialized Search cont'd/ Final Exam		

- This schedule, and the topics here, may change as the course progresses.
  The schedule, readings and deadlines will be updated appropriately.
- By default, all readings are from:
  Search Engines: Information Retrieval in Practice.
  Bruce Croft, Donald Metzler and Trevor Strohman,
  Addison-Wesley. 2010. ISBN: 978-0-136-07224-9 [Abbreviated as CMS]
- There will be a quiz (almost) each week, and quiz responses will be due by 9am Pacific of the next Monday. Homework Assignments are also due at 9am Pacific on the specified day.